

SentiVision

Sentimentalists

Horváth István, Kónya Leon

Felkészítő tanár: Kőrösi Gábor

*Bolyai Tehetséggondozó Gimnázium és Kollégium,
24400 Zenta, Posta utca 18*

1. Bevezetés

Célunk egy olyan eszköz kifejlesztése, elkészítése volt, amely képfeldolgozás segítségével könnyíti a vakok illetve látássérültek mindennapi funkcióit, konkrétan az érzelem felismerést. Egy olyan mobilapplikáció volt a vízióink, amely a felhasználó társalgási partnerének arckifejezéseit elemzi ki valós időben, és annak pillanatnyi érzelméről adott becslését közli a felhasználóval nem zavaró hanghatás formájában.

A vízióink ötletét olyan kísérletek adták, amelyek a vak vagy látássérült illető által viselt kamera felvételeit próbálták hang formájába önteni, hogy azok segítsék őt mindennapi tevékenységei hátránytalan végzésében. Viszont ezek nem arattak nagyobb sikert, és a képfeldolgozó algoritmusaik aránylag primitívek voltak.

Az utóbbi évtizedben azonban robbanásszerűen növekedni kezdett a GPU-ipar, így elterjedtebbek lettek a nyílt forráskódú, neurális háló alapú gépi látási algoritmusok. Ennek előnyét számos tech-óriás kihasználja, kommerciális szinten és ipari szinten is (pl. önvezető autók).

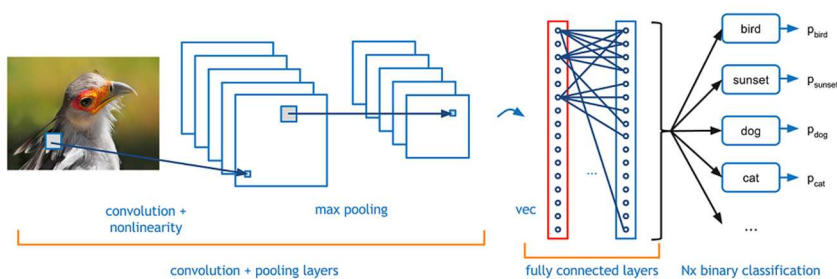
Ez a két gondolatmenet összekombinálása adta számunka az ötletet, ugyanis nem találkoztunk még vakokat segítő konvolúciós neurális hálókkal.

2. Probléma megoldásának menete

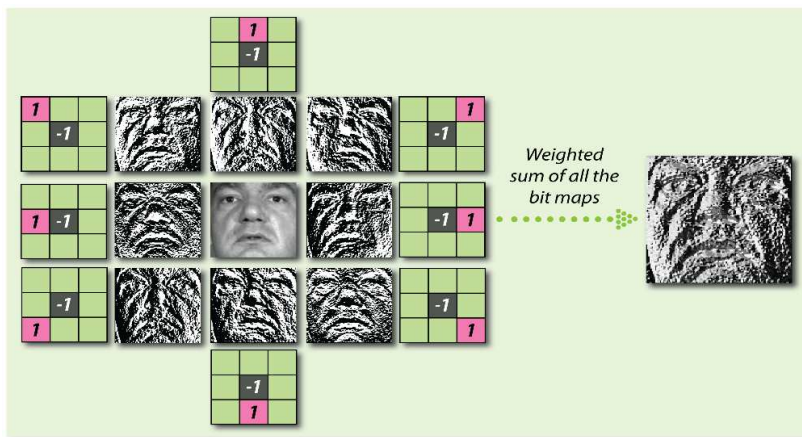
A munkánk kivitelezése a következő főbb problémák megoldását igényelte: az arcelemző algoritmusunk kidolgozása során értelmetlennek láttuk a 0-ról indulást, így egy meglévő, előre betanított CNN (Convolutional Neural Net) architektúrát kellett találnunk, amely elég kompakt ahhoz, hogy egy okostelefon futtatni tudja valós időben. A MobileNetre esett a választásunk. Ezek után egy hasonlóan kompakt arckifejezés-adatbázist kellett találnunk, amelyen az eredetileg ImageNetre tanított MobileNetet „újratanítjuk”. Ez a FER2013-as adatbázis lett. Python környezetben, Tensorflow könyvtár segítségével lett a feladat elvégezve. Az applikáció Firebase platform segítségével lett létrehozva.

2.1. Konvolúciós neurális hálók az arcfelismerésben

A konvolúciós neurális hálók működési elve nagyvonalakban az 1. ábrán látható. Tradicionális neurális hálókkal szemben előnyös képfeldolgozásnál, mivel úgynevezett „konvolúciós kerneleket” használ: $n \times n$ -es súlyozott mátrixokat, amelyeknek a bemeneti kép minden $n \times n$ -es szegmensének és a kernelnek skaláris szorzata által alkotott feldolgozott képet adják tovább a hálózat további rétegeinek. A 2. ábra szemlélteti, hogy mi is történik egy konvolúciós művelet során.



1. ábra: A konvolúciós hálózatok



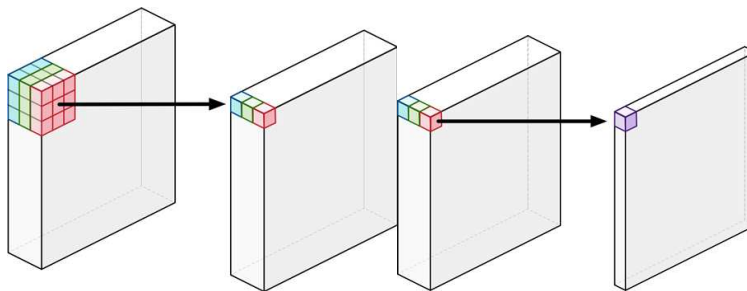
2. ábra: Különböző konvolúciós kernelek

Látható, hogy különbözőképpen súlyozott kernelek különböző tulajdonságokat emelnek ki egy képen. Azt pedig, hogy mely tulajdonságok lesznek kiemelve, maga a hálózat alakítja ki, az alapján, hogy mi a leghatékosabb neki a tanulási folyamat során, példák alapján.

Mint ahogyan a sztandard neurális hálózatoknál, úgy a konvolúciósaknál is gradiens süllyedéssel illetve hátrapropagációval történik a tanulás. Ez gyakorlatilag kategóriába való osztályozás során egy hibafüggvény definiálásával kezdődik a hálózat összes paraméterének függvényében, amelynek kimenetelét minimalizálja Newton numerikus módszereivel, és ez alapján változtatja a hálózat összes paraméterét a hibafüggvény minimuma felé.

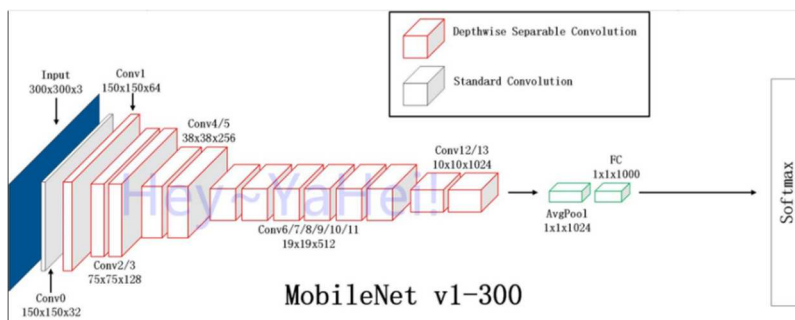
2.2. Mobilenet

A Mobilenet tökéletes gépi látási architektúra mobilos alkalmazásokhoz, ugyanis kisebb kapacitású számításoknál ugyanannyi paramétermennyiség esetén (Mobilenet-nél 4.24 millió) az eredeti teljesítménynek többszörösét is el tudja érni egy pár egyszerű trükkel, köztük a mélységi és a pontbeli konvolúciók szétválasztásával (a 3. dimenzió a színsatorna), amint az a 3. ábrán is látható.



3. ábra: A két konvolúció szétválasztása

Maga a Mobilenet architektúrája a 4. ábrán látható.



4. ábra: Mobilnet architektúrája

A hálózat eredetileg egy sokkategóriás képosztályozó kategóriára lett betanítva, amely ismét ideális nekünk, ugyanis feltételezhetjük, hogy sok általános tulajdonságot elsajátított a tanulás során (ImageNet konvolúciós súlyok), amelyeket újratanításkor csak nagyon minimálisan kell módosítani, hogy jól tudjon osztályozni.

2.3. Újratanítás után

Újratanítás után elkészítettük a confusion matrix-et, amely 0-tól 3-ig indexelt soraiba és oszlopaiba az undor, neutrális, mosolygó illetve meglepődött arckifejezéseket kell behelyettesíteni. Ebben a mátrixban (5. ábra) leolvasható, hogy mely arckifejezéseket mekkora számban tévesztett a hálózat össze teszteléskor melyekkel.

```
Confusion Matrix
[[ 1   2  45   7]
 [ 1  69 230 326]
 [ 22 463 288 106]
 [ 0  25 342  49]]
```

5. ábra: Confusion Matrix az újratanított Mobilenethez

Első látásra nem tűnik úgy, hogy nagy precizitása lenne a hálózatnak, viszont az applikáció működésben rezponzív legalább 3 arckifejezésre. Nagyobb, professzionálisabb adatbázisokkal ez orvosolható, a célunk egyelőre a rezponzív működés volt.

3. Elért eredmények

Eddigi eredményeink arra utalnak, hogy az általunk elképzelt eszköz kivitelezhető, életképes. Elegendő adattal számos hiányosság pótolható, ezért a jövőben szert szeretnénk tenni akadémiai adatbázisokra, több arckifejezés bevonásával.

Az eszközben főleg azért látunk potenciált, mert nem verseng a vakvezető kutyával, illetve a bottal sem. Ezeket évtizedek óta próbálják szakemberek lecserélni, korszerűsíteni, ám nem lehet őket legyőzni. Éppen ezért a SentiVision ezeket kiegészíteni kívánja, egy olyan funkciót ad az ember kezébe, amelyet a bot és kutya nem tudnak megadni.

Ennek az innovatív funkciónak köszönhetően feltételezni lehet, hogy termék formájában lenne piaca az erre igényt tartók körében. Számtalanszor könnyebb lehet egy konverzáció, hogyha a vak ember tisztában van a társalgási partnere nonverbális jeleivel – végülis a testbeszéd többet mond minden szónál.